Improving Assessment and Instruction with Process Data and Learning analytics

Mingyu Feng

Senior Research Associate, WestEd

Nov 7, 2019 2019 MARC Conference

Overview

Two strands of work we have done using learning analytics based on continuous process data from technology-based learning systems

1. Assess Student Learning Outcomes
2. Support Teacher Practices and Improve Student Learning

SECTION 1

Using Process Data to Assess Learning Outcome

In ASSISTments

Motivation

Concerns about poor student performance on new state tests

- High-stakes standards-based tests are required every state
- Student performance are not satisfactory
  - **National report card 2017, only 34% percent of 8th students performed at or above the *Proficient* level on the mathematics assessment**

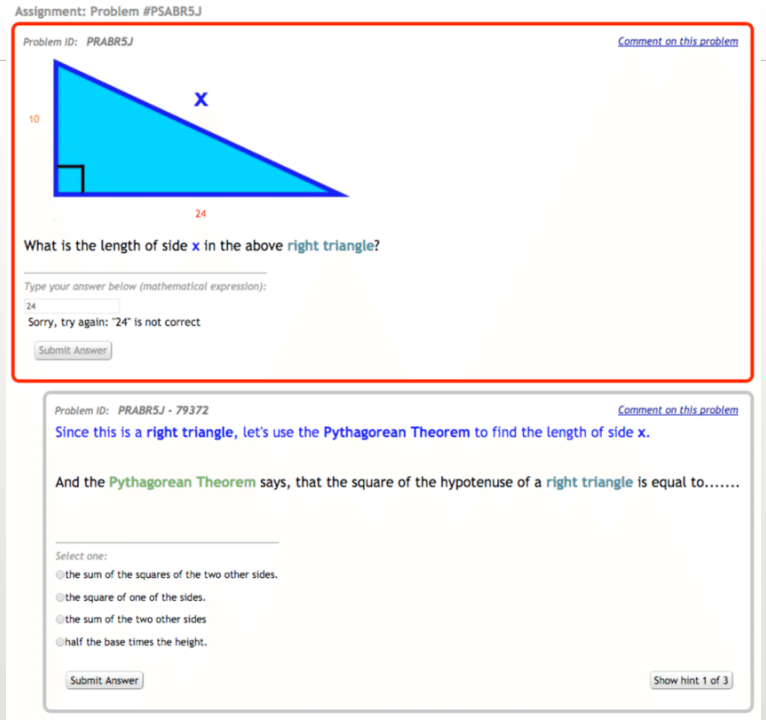
Secondary teachers are asked to be data-driven

- Assessment takes valuable classroom time away
  - **Can we assess student performance without sufficing learning time?**
- Not all kinds of data are equally accessible and informative
  - **State test reports, even interim assessment reports are not readily available soon enough to guide instruction**
  - **The integration of DDI takes great effort of teachers**

ASSISTments: Blending Assisting and Assessment

When solving problems in ASSISTments,

- Students get immediate feedback, hints, guided practice / scaffolding questions



ASSISTments™

A Free Public Service of Worcester Polytechnic Institute

A Grade Book Report

Student Name	Elapsed time (hh:mm)	Original Items				Scaffolding + Original Items		
		# Done	% Correct	Est. MCAS Scaled Score*	Est.MCAS Performance Level	# Done	% Correct	# Hint Req.
Tom*	4:12	90	39%	214	Warning/Failing	228	44%	353
Dick*	4:01	98	66%	244	Proficient	158	59%	58
Harry*	4:07	78	40%	224	Needs improvement	154	38%	77
Mary*	4:17	114	20%	200	Warning/Failing	356	20%	705
Jack*	3:53	104	39%	214	Warning/Failing	267	43%	187
				244	Proficient	40	52%	55

Where does this score come from?

Why should we be more complicated?

Student Name	Elapsed time (hh:mm)	Original Items				Scaffolding + Original Items		
		# Done	% Correct	Est. MCAS Scaled Score*	Est.MCAS Performance Level	# Done	% Correct	# Hint Req.
Tom*	4:12	90	39%	214	Warning/Failing	228	44%	353
Dick*	4:01	98	66%	244	Proficient	158	59%	58
Harry*	4:07	78	40%	224	Needs improvement	154	38%	77
Mary*	4:17	114	20%	200	Warning/Failing	356	20%	705
Jack*	3:53	104	39%	214	Warning/Failing	267	43%	187
John*	4:24	92	55%	244	Proficient	40	52%	55

- Static – does not distinguish "Tom" and "Jack" → **Dynamic assessment**
- Average – ignores development over time → **Longitudinal modeling**
- Uninformative – not informative for classroom instruction → **Cognitive diagnostic assessment**

7

Dynamic Assessment – The idea

Student Name	Elapsed time (hh:mm)	Original Items				Scaffolding + Original Items		
		# Done	% Correct	Est. MCAS Scaled Score*	Est.MCAS Performance Level	# Done	% Correct	# Hint Req.
Tom*	4:12	90	39%	214	Warning/Failing	228	44%	353
Dick*	4:01	98	66%	244	Proficient	158	59%	58
Harry*	4:07	78	40%	224	Needs improvement	154	38%	77
Mary*	4:17	114	20%	200	Warning/Failing	356	20%	705
Jack*	3:53	104	39%	214	Warning/Failing	267	43%	187
John*	4:24	92	55%	244	Proficient	40	52%	55

- Dynamic testing began before computerized testing (Brown, Bryant, & Campione, 1983).

8
8

Dynamic vs. Static Assessment

Dynamic testing metrics

- # attempts
- # minutes to come up with an answer; # minutes to complete an ASSISTments
- # hint requests; # hint-before-attempt requests; #bottom-out hints
- % correct on scaffolds
- # problems solved

"Static" measure

- correct/wrong on original questions

9
9

Dynamic Assessment - Modeling

Three linear stepwise regression models

1-parameter IRT proficiency estimate

The standard test model

MCAS Score

1-parameter IRT proficiency estimate + all online metrics

The mixed model

The assistance model

All online metrics

10

10

Dynamic Assessment - Evaluation

Bayesian Information Criterion (BIC)

$$BIC = -2\log(L) + p\log(n)$$

- Widely used model selection criterion
- Resolves overfitting problem by introducing a penalty term for the number of parameters
- Formula
- Prefer model with lower BIC

Mean Absolute Deviation (MAD)

- Cross-validated prediction error

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|MCAS_i - pred_i|$$

$$Pct\_Err = MAD/(Max\ Raw\ Score)$$

- Prefer model with lower MAD

11

11

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.

Dynamic Assessment - Results

1-parameter IRT proficiency estimate

The standard test model

MCAS Score

1-parameter IRT proficiency estimate + all online metrics

The mixed model

The assistance model

All online metrics

Model	MAD	BIC	Correlation with 2005 8th grade MCAS
The standard test model	6.40	-295	0.733
The assistance model	5.46	-402	0.821
The mixed model	5.04	-450	0.841

p=0.001

p=0.001

Dynamic Assessment - Robustness

- See if model can generalize
  - Train/test model on split of same year's data
  - Test model on other year's data

Results of testing the mixed models on a different year's data

Year of training data	Year of Testing data	MAD on Testing data	Correlation with MCAS scores on testing data
2004-2005	2005-2006	5.16	0.827
2005-2006	2004-2005	5.80	0.824

Section 1: Conclusion

ASSISTments process data enables us to assess more accurately

The relative success of the assistance model over the standard test model highlights the power of the dynamic measures

14

SECTION 2

Leveraging Learning Analytics to Support Teacher Practices and Improve Student Learning

In ASSISTments

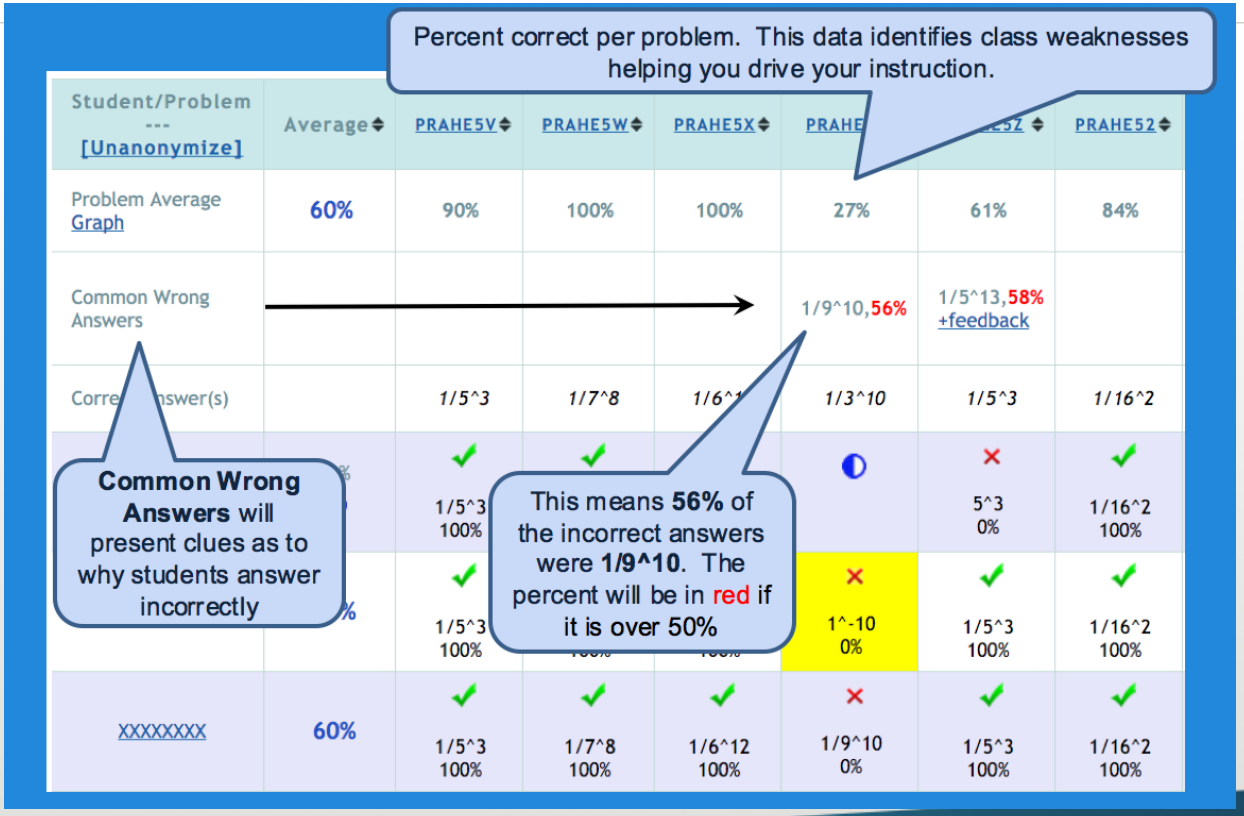
ASSISTments for homework support

- Students get immediate feedback, hints, guided practice while doing homework
- Teachers can see how students are doing on homework
- Teachers can change homework reviews to adapt to students

ASSISTments™

A Free Public Service of Worcester Polytechnic Institute

ASSISTments Item Report



Student/Problem --- [Unanonymize]	Average	PRAHE5V	PRAHE5W	PRAHE5X	PRAHE	PRAHE52
Problem Average Graph	60%	90%	100%	100%	27%	61%	84%
Common Wrong Answers					1/9^10,56%	1/5^13,58% +feedback	
Corre nswer(s)		1/5^3	1/7^8	1/6^1	1/3^10	1/5^3	1/16^2
		✓ 1/5^3 100%	✓		◐	✗ 5^3 0%	✓ 1/16^2 100%
		✓ 1/5^3 100%	✓		✗ 1^-10 0%	✓ 1/5^3 100%	✓ 1/16^2 100%
XXXXXXXX	60%	✓ 1/5^3 100%	✓ 1/7^8 100%	✓ 1/6^12 100%	✗ 1/9^10 0%	✓ 1/5^3 100%	✓ 1/16^2 100%

Research Overview

A Goal 3 Efficacy Trial

Sample

- Maine gives all students a laptop to take home, hence hardware is available
- Focused on 7th grade math students
- Recruited two cohorts:
  - Summer 2012 - June 2014
  - Summer 2013 - June 2015
- 46 schools recruited
- 43 schools in final sample
  - 87 teachers
  - 2800+ students

Research Overview

A Goal 3 Efficacy Trial

Random Assignment

- Schools paired by similar size and prior math scores
- “Coin toss” for each pair

Schools stay in condition for two years

- Year 1: Teacher preparation, practice and coaching
- Year 2: Teachers use with new cohort of students

Teachers expected to follow school homework policies

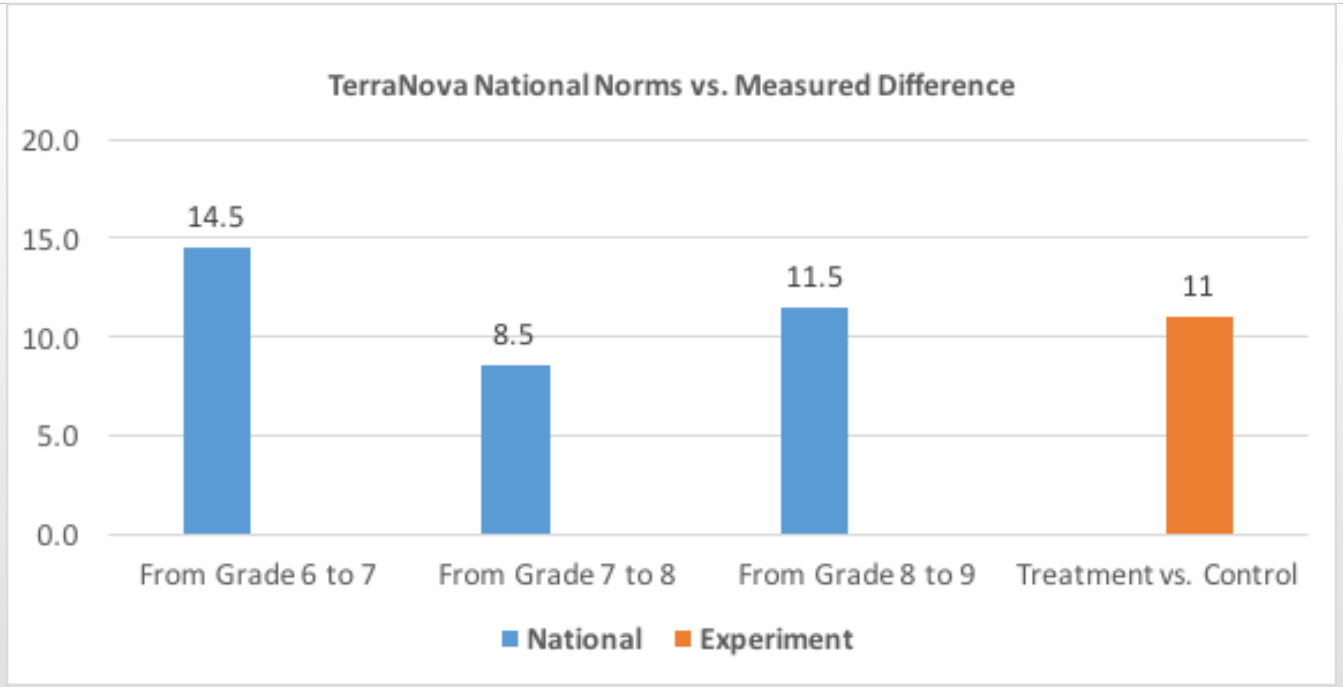
Content comes from their textbooks, plus pre-built content in ASSISTments

Findings

1. Impact on student learning
2. Benefit to students with low prior math achievement
3. Change in teacher practices

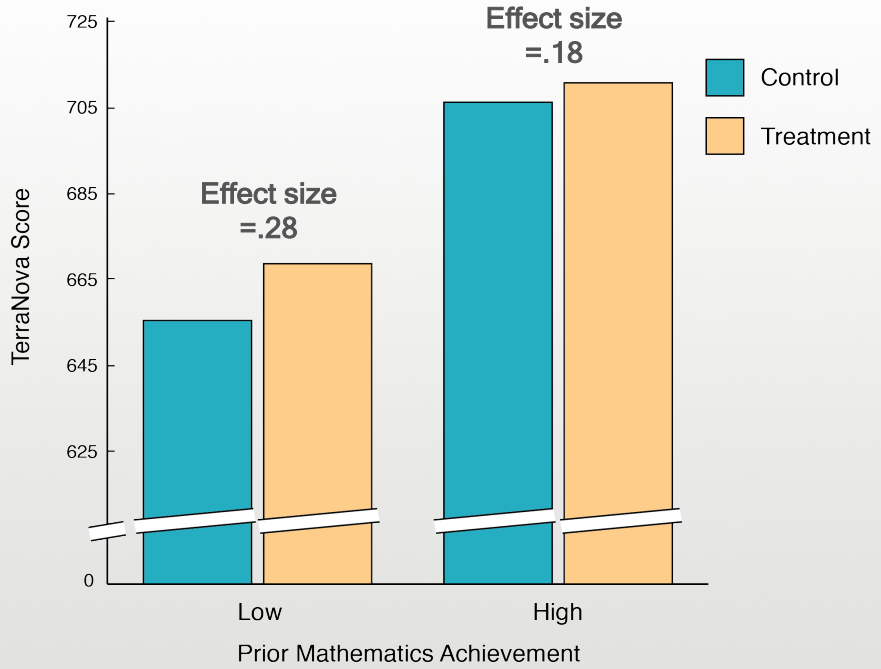
Finding 1: Impacts on Student Achievement

effect size = 0.18; Unadjusted TerraNova Score 11 points higher



Finding 2: Greater Effect for Low Prior Math

Via Median Split Analysis



Makes sense: higher achievers need less homework help

Finding 3: Change in Teacher Practice

- Shift from reviewing "all problems" to targeted problems

"Now it's easier for me to go through and pick out the specific ones that I want them to do. . . . I can use the reports to target those questions, those areas that it's obvious that the entire class or most of the class is struggling with. So I can go back and spend a little more time with that."

How Did You Determine Which Problems to Review? (Survey)			
	Treatment	Control	P Value
ALL the problems	24%	43%	0.000
Targeted problems	75%	36%	0.000

How Did You Determine Which Problems to Review? (Log)			
	Treatment	Control	P Value
All the problems	21%	21%	0.968
Targeted Problems	93%	68%	0.003
Student Driven	57%	76%	0.072

Finding 3: Change in Teacher Practice

- Homework review: same total time, fewer problems
- Emphasis on problems that were difficult for students
- Addressing student errors

"I'm not having to go over every single problem. I can look at the big picture and say, 'All right, this is how you did.' Kids see it, I see it . . . you know, 'Only 30 percent of the class got this right. We're definitely looking at this question.'"

"We do more talking and discussing about the homework. . . . not just finding and fixing, but they're able to discuss maybe a process or a step that they're missing. . . . I think it's really changed how we talk about homework, as opposed to right and wrong answers. But what we got wrong and why."

Section 2: Conclusion

Homework CAN be improved

- Intervention needs to be easy for the public to digest – they care about homework
- Relatively easy for schools to adopt –modest intervention in an area of stable school
- *How* to do homework, not what or how much

Teachers' practice CAN be changed

- When the intervention can be readily integrated into general practices
- When the right kind of data is available
- There is support on strategies to respond to the data

More On-going Studies

Replication study in North Carolina – IES funded, led by WestEd

- Face-to-face training + local coach visits

National effectiveness study – IES funded, led by AIR

- Face-to-face training + Google hang-out for support

Scale up study – DoE EIR, led by WPI

- All virtual training + Instructional recommendation based on common errors

Thanks!

Questions?

mfeng@wested.org

Summary

The data we focused on is

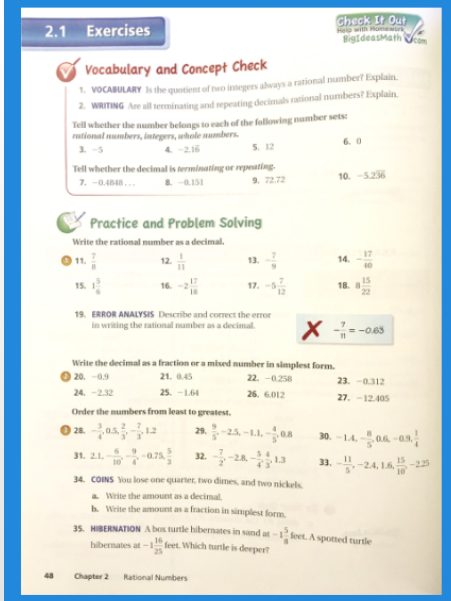
- Individualized
- Generated by online *learning* systems
- Unobtrusive collected at scale
- Large amount (and noisy)
- Behavior or performance data

Best practices

- Identify useful data
- Become familiar with teachers' and students' experience of using the online platforms
- Understand the ways in which these data were collected and stored by the system
- Build explainable, actionable analytics that need low-effort for adoption and integration
- Check face validity and external validity of measures

ASSISTments Used for Textbook Work

Students use their textbook or worksheet and...



2.1 Exercises

Vocabulary and Concept Check

1. VOCABULARY Is the quotient of two integers always a rational number? Explain.
2. WRITING Are all terminating and repeating decimals rational numbers? Explain.

Tell whether the number belongs to each of the following number sets: *rational numbers, integers, whole numbers.*

3. -5 4. -2.16 5. 12 6. 0

Tell whether the decimal is *terminating* or *repeating*.

7. $-0.4848\ldots$ 8. -0.151 9. 72.72 10. -5.236

Practice and Problem Solving

Write the rational number as a decimal.

11. $\frac{7}{8}$ 12. $\frac{1}{11}$ 13. $-\frac{7}{9}$ 14. $-\frac{17}{40}$
15. $1\frac{5}{8}$ 16. $-2\frac{17}{18}$ 17. $-5\frac{7}{12}$ 18. $8\frac{15}{22}$

19. ERROR ANALYSIS Describe and correct the error in writing the rational number as a decimal. $\frac{7}{11} = 0.65$

Write the decimal as a fraction or a mixed number in simplest form.

20. -0.9 21. 0.45 22. -0.258 23. -0.312
24. -2.32 25. -1.64 26. 6.012 27. -12.405

Order the numbers from least to greatest.

28. $-\frac{3}{4}, 0.5, \frac{2}{3}, -\frac{7}{3}, 1.2$ 29. $\frac{9}{5}, -2.5, -1.1, -\frac{4}{5}, 0.8$ 30. $-1.4, -\frac{8}{5}, 0.6, -0.9, \frac{1}{4}$
31. $2.1, -\frac{6}{10}, -\frac{9}{4}, -0.75, \frac{5}{2}$ 32. $-\frac{7}{2}, -2.8, -\frac{5}{4}, \frac{4}{3}, 1.3$ 33. $-\frac{11}{5}, -2.4, 1.6, \frac{15}{10}, -2.25$

34. COINS You lose one quarter, two dimes, and two nickels.
a. Write the amount as a decimal.
b. Write the amount as a fraction in simplest form.

35. HIBERNATION A box turtle hibernates in sand at $-1\frac{5}{8}$ feet. A spotted turtle hibernates at $-1\frac{16}{25}$ feet. Which turtle is deeper?

48 Chapter 2 Rational Numbers

...do their work on paper. After completing each problem...

...they enter their answers into ASSISTments to learn whether they are correct or incorrect.

The students are learning from their mistakes and have a better idea of areas in which they need additional help.

Julie Schoenfeld, 7th Grade Math Teacher, Sudbury, MA

The students benefit by knowing when they are doing a problem wrong allowing them to fix their approach before moving on to the next problem instead of continuing to do the work wrong.

Vicki Davis, 7th Grade Math Teacher, Blue Hills, ME

Design Workshops - where data became actionable

Measures were put in front of faculty to better understand

- how well they captured students' learning strategies and behaviors
- whether they could be used to help faculty more effectively intervene with students

Researchers + Faculty

- Researchers learn from faculty on what they found meaningful and whether certain patterns that were identified had face validity
- An efficient touch-point for faculty to engage in data-intensive research activities

Structured activities

- Jointly interpreting data products
- co-developing data products and follow up actions such as change ideas that faculty could implement using a data